# Bayesian Analysis of Mass Spec Enrichment in integrative studies of chromatin-associated complexes

Peter V. Kharchenko, Artyom A. Alekseyenko, Andrey A. Gorchakov, Mitzi I. Kuroda
*Harvard Medical School, Boston, MA,* peter.kharchenko@post.harvard.edu

Chromatin composition is intricately linked with the transcriptional and regulatory activity of the associated genomic loci. To understand the molecular mechanisms underlying these processes, it is important to obtain a comprehensive picture of the chromatin protein, DNA, and RNA components, as well as their mutual interactions. Towards that aim we have developed a high-stringency cross-linking approach allowing for mass-spectrometry identification of protein-protein interactions together with high-resolution mapping of RNA- and DNA-protein interactions. We examine chromatin composition associated with two contrasting *D. melanogaster* proteins, MSL3 and HP1a, known to have a distinct repertoire of RNA and protein interactions, and localizing to euchromatic and heterochromatic compartments respectively.

Here we focus on the Bayesian model for analysis of associated protein content. The chromatin fraction enriched for a protein of interest (*i.e.* MSL3) is analyzed using liquid chromatography–mass spectrometry (LC-MS). Each individual LC-MS experiment can be summarized by unique peptide counts – the number of distinct fragments encountered for each protein in a given sample. Such peptide counts, however, depend on a variety of factors, including the length of the protein, its aminoacid composition, and the total protein load in a given LC-MS run. A quantitative assessment of unique peptide counts therefore requires elaborate computational procedures [1, 2].

To identify proteins that show significant association with MSL3 or another protein of interest, we developed a flexible statistical approach that controls for multiple sources of bias, including replicate variability, differential base-level abundance of proteins in different cell types, as well

as non-specific effects of pull-down procedures. In evaluating whether a given protein shows MSL3-specific enrichment, the method uses a hierarchical Bayesian model to examine the likelihood of possible values for three key parameters: base-level abundance of that protein in each cell type; the extent to which the protein is enriched due to mock or non-specific IPs; and the extent to which the protein is enriched by the specific (*i.e.* MSL3) IP. The probability that a given protein shows enrichment magnitude in the MSL3 pull down given the observed LC-MS data, is summarized by the posterior distribution of  that takes into account possible contributions of other factors and the variability observed between replicates.

Our analysis of MSL3 and HP1a-enriched chromatin fractions recovers multiple known and novel interaction partners, along with associated non-coding RNAs, and detailed genomic occupancy profile. We believe our approach can serve as a robust tool for characterization of chromatin composition.

References:


1 Choi, H. *et al.* SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* 2011 Jan;8(1):70-3

2 Sowa M.E. *et al.* Defining the human deubiquitinating enzyme interaction landscape. *Cell* 2009 Jul 23;138(2):389